

### REMARKS

In response to the Office Action mailed August 30, 2011, the Assignee of the present application (*Nuance Communications, Inc.*) respectfully requests reconsideration. Claims 1, 4, 5, 7-11, 14, 15, 17-23 and 29-35 were previously pending for examination. By this amendment, claims 22 and 23 are canceled without prejudice or disclaimer. Claims 1, 4, 5, 7-11, 17-19, 21, 29, 30 and 33-35 have been amended herein. Claims 36 and 37 have been added. As a result, claims 1, 4, 5, 7-11, 14, 15, 17-21 and 29-37 are currently pending for examination, with claims 1, 11 and 21 being independent. No new matter has been added.

The claim amendments and newly added claims are supported throughout the specification. For example, support for the amendments common to independent claims 1, 11 and 21 can be found in the specification at least at page 9, lines 9-12, and page 9, line 22 – page 10, line 8. Support for newly added claims 36 and 37 can be found at least at page 16, line 9 – page 17, line 16.

### Rejections Under 35 U.S.C. § 101

The Office Action rejects claims 1, 4, 5, 7-10, 21-23, 29, 30, 33 and 35 under 35 U.S.C. § 101 as purportedly being directed to non-statutory subject matter.

At page 2, the Office Action asserts that claims 1, 4, 5, 7-10, 29, 30 and 33 “fail[] to be limited to only statutory embodiments” because they allegedly are not “limited to non-transitory storage.” Without acceding to the propriety of this rejection, the Assignee has amended each of claims 1, 4, 5, 7-10, 29, 30 and 33 herein to recite an “article of manufacture” comprising a program storage device. Such an article of manufacture is statutory subject matter under 35 U.S.C. § 101, which states that “[w]hoever invents or discovers any new and useful process, machine, **manufacture**, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor” (emphasis added). Accordingly, withdrawal of the rejection of claims 1, 4, 5, 7-10, 29, 30 and 33 under 35 U.S.C. § 101 is respectfully requested.

At page 2, the Office Action asserts that claims 21-23 and 35 “fail to recite any physical hardware or elements of an actual system.” Without acceding to the propriety of this rejection, the Assignee has amended independent claim 21 herein to recite a system comprising “at least one processor; and at least one storage device.” Accordingly, withdrawal of the rejection of claims 21-23 and 35 under 35 U.S.C. § 101 is respectfully requested.

### Rejections over the Cited Art

The Office Action rejects each of the independent claims (i.e., claims 1, 11 and 21) under 35 U.S.C. § 102(b) as purportedly being anticipated by Eide (U.S. Patent No. 6,101,470). The Office Action rejects each of the dependent claims either under 35 U.S.C. § 102(b) as purportedly being anticipated by Eide, or under 35 U.S.C. § 103(a) as purportedly being obvious over Eide in view of one or more secondary references. The Assignee respectfully traverses these rejections.

#### I. Discussion of Some Embodiments

Some embodiments described in the present application relate to converting a text input to synthesized speech in a manner that mimics the style and pronunciation of a spoken example of the text input (page 1, lines 6-10). When a user supplies a spoken example of a text string to be text-to-speech synthesized, some embodiments can extract prosodic parameters from the spoken example, and adopt those prosodic parameters as synthesis parameters for generating the synthetic speech waveform (page 8, lines 1-8). Prosodic parameters include speech quality attributes, such as pitch, duration, energy values, etc. for multiple speech segments in the spoken audio signal (page 1, lines 21-42; page 10, lines 23-24).

Some embodiments provide a user interface that allows a user to identify a text string and speak an example of a desired pronunciation of the text string (page 9, lines 9-12). A prosody analyzer can then process the spoken audio signal to extract prosodic parameters from it, such as pitch and energy contours (i.e., sets of pitch and energy values extracted from various time indexes during the audio signal) (page 10, line 20 – page 11, line 1). An alignment module can extract the

duration of each unit of the text (e.g., word or phoneme) as produced in the spoken example (page 13, lines 18-21). This can be accomplished by aligning the audio signal with the text.

After prosodic parameters including durational parameters have been extracted by the prosody analyzer and adopted as synthesis parameters, a conversion module can process the synthesis parameters to generate an input for a text-to-speech (TTS) engine (page 15, line 24 – page 16, line 4). The input may be in a format such as SSML (Speech Synthesis Markup Language), which provides the text string with prosody markup that can be processed by the TTS engine to produce a synthetic speech waveform with the specified prosodic parameters (page 16, line 9 – page 17, line 7). The duration parameters extracted by the alignment module may allow the other prosodic parameters from various time indexes during the audio signal to be mapped to the appropriate text units when the text is converted to synthetic speech (page 16, lines 5-8). The TTS input so generated may then be used by the TTS engine in converting the text string to synthetic speech (page 18, lines 9-11).

The foregoing discussion is provided to assist the Examiner in appreciating some aspects of the invention. However, this discussion may not apply to each of the independent claims, and the language of each independent claim may differ in material respects from the discussion provided above. Therefore, the Assignee respectfully requests that careful consideration be given to the language of each independent claim, and that each be addressed on its own merits, without relying on the discussion above. In this respect, the Assignee does not rely on the discussion above to distinguish any of the claims over the cited art, but rather relies only upon the claim language and the arguments presented below.

## II. Discussion of Eide

Eide relates to a method for automatically generating pitch contours in a text to speech (TTS) system. Eide: Abstract. The system uses a pitch model that is trained on a reading of a training text. Eide: col. 4, lines 20-25. A training speaker reads the training text, and the pitch of the training speaker's voice is measured as the speaker speaks. Eide: col. 5, line 54 – col. 6, line 10.

Using a phonetic dictionary, the level of lexical stress of each vowel in the training text is determined. Eide: col. 6, lines 35-49. By matching the stress level of each vowel in the training text with the measured pitch of the training speaker's voice when the training speaker read that vowel, a pitch value is paired with a lexical stress level for each syllable in the training text. Eide: col. 6, lines 50-55. These (lexical stress, pitch) pairs are then stored to form the pitch model. Eide: col. 6, lines 64-66.

When the pitch model is to be used in TTS synthesis, an input text to be synthesized is obtained. Eide: col. 4, lines 32-34. Using the phonetic dictionary, the stress contour of each input sentence is calculated, by determining the lexical stress level of each vowel in the input text. Eide: col. 4, lines 35-45. The stress contour of the input sentence is then divided into blocks, and the training text used to train the pitch model is searched for matching stress contour blocks. Eide: col. 5, lines 8-25. For each block of the input sentence having a stress contour, a block of the training text having a matching stress contour is selected. Eide: col. 5, lines 26-32. This results in a sequence of training blocks in which each of the training blocks may have come from a different sentence in the training text. Eide: col. 9, line 62 – col. 10, line 12. The measured pitch values (as spoken by the training speaker) associated with the selected training blocks are then concatenated to create a pitch contour for use in synthesizing the input text to speech. Eide: col. 5, lines 33-40.

### III. The Claims Patentably Distinguish over Eide.

Independent claim 11 is directed to a method comprising, *inter alia*, "providing a user interface that allows a user to identify a text string for synthesis and to speak a pronunciation of the text string; recording the user's spoken pronunciation of the text string as an audio signal;... adopting as synthesis parameter values the prosodic parameter values extracted from the audio signal; and generating a synthetic speech waveform representing the text string using the synthesis parameter values." Independent claim 1 recites "[a]n article of manufacture comprising a program storage device... tangibly embodying a program of instructions executable..." to perform the same method. Independent claim 21 recites "[a] text-to-speech (TTS) system... comprising: at least one

processor; and at least one storage device storing processor-executable instructions that..." perform the same method. Eide fails to describe such a method for generating synthetic speech.

As discussed above, Eide trains a pitch model by having a training speaker read a training text, measuring the pitch of the training speaker's voice while reading the training text, and collating the measured pitch with the stress contour of the training text. Eide: col. 5, line 54 – col. 6, line 66. The Office Action (page 3), in alleging that Eide teaches extracting prosodic parameter values from an audio signal corresponding to a pronunciation of a text string by a user, cites to the section of Eide about training the pitch model. The Office Action thus appears to equate Eide's reading of the training text by the training speaker with the user's pronunciation of the text string as claimed in the present application. However, in the claims of the present application, the text string that is identified by the user is a text string *for synthesis*, and is the same text string represented by the synthetic speech waveform that is generated. Eide's training text does not meet these limitations, as Eide's training text is not the text that is synthesized. Eide's training text is simply a text that is read by a training speaker to train the pitch model; no mention is made of generating any synthetic speech waveform representing the training text. The input text that is synthesized in Eide is different from the training text used to create the pitch model.

In addition, Eide does not provide "a user interface that allows a user to identify a text string for synthesis and to speak a pronunciation of the text string," as recited in the independent claims. In Eide, the user enters an input text to be synthesized (Eide: col. 7, lines 40-41), but no mention is made of allowing the user to speak a pronunciation of that input text. Eide thus does not allow a user-specified pronunciation in the manner recited in the independent claims.

For at least these reasons, each of independent claims 1, 11 and 21 patentably distinguishes over Eide, and withdrawal of the rejections of these claims is respectfully requested.

Each of claims 4, 5, 7-10, 14, 15, 17-20 and 29-35 depends from an independent claim discussed above and is allowable for at least the same reasons. Accordingly, withdrawal of the rejections of these claims is respectfully requested.

Each of newly added claims 36 and 37 also depends from an independent claim discussed above and is allowable based at least on its dependency. Accordingly, allowance of these claims is respectfully requested.

#### General Comments on Dependent Claims

Because each of the dependent claims depends from a base claim that is believed to be in condition for allowance, the Assignee believes that it is unnecessary at this time to argue the further distinguishing features of all of the dependent claims. However, the Assignee does not necessarily concur with the interpretation of the dependent claims as set forth in the Office Action, nor does the Assignee concur that the basis for the rejection of any of the dependent claims is proper. Therefore, the Assignee reserves the right to specifically address in the future the further patentability of the dependent claims not specifically addressed herein.

**CONCLUSION**

In view of the foregoing, the present application is believed to be in condition for allowance. A Notice of Allowance is respectfully requested. The Examiner is requested to call the undersigned at the telephone number listed below if this communication does not place the case in condition for allowance to discuss any outstanding issues relating to the allowability of this application.

If the response is not considered timely filed and if a request for an extension of time is otherwise absent, the Assignee hereby requests any necessary extension of time. The Assignee believes no fee is due with this response. However, if a fee is due, please charge Deposit Account No. 23/2825 under Docket No. N0484.70760US00 from which the undersigned is authorized to draw.

Dated: November 30, 2011

Respectfully submitted,  
Nuance Communications, Inc.

By: 

Richard F. Giunta  
Registration No.: 36,149  
Elisabeth H. Hunt  
Registration No.: 67,336  
WOLF, GREENFIELD & SACKS, P.C.  
Federal Reserve Plaza  
600 Atlantic Avenue  
Boston, Massachusetts 02210-2206  
617.646.8000